[Email address]

Report

August 2024

Contents

[Introduction and Background: 3](#_Toc174906780)

[Collaboration: 4](#_Toc174906781)

## Introduction and Background:

In the dynamic landscape of machine learning, classification and image recognition occupy a position of paramount importance. This report highlights and exploration of a simple classification problem and a mutli-label image-based digit classification. Delving into the intricacies of classification, its pivotal role in unravelling patterns within data and the creation of machine learning models creating the world of AI as we currently know.

***The Significance of Classification in Machine Learning:***

Classification, at its core, is the task of assigning predefined labels or categories to instances based on their inherent characteristics. It serves as the backbone of numerous real-word applications, ranging from spam email detection and sentiment analysis to medical diagnosis and fraud prevention. The ability to accurately categorize data empowers machines to make informed decisions, automate processes, and extract valuable insights that drive innovation and progress across diverse domains.

***Project purpose and objectives:***

The focal point of this project is to harness the power of machine and deep learning algorithms to tackle a specific classification problem. By meticulously designing, training, and evaluating models, we aim to achieve exceptional classification performance, characterized by high accuracy, precision, and recall. The overarching objectives include:

1. **Comprehensive Data Analysis and Visualisation:** Gain a deep understanding of the dataset, identifying relevant features, addressing imbalances, and preprocessing data for optimal model performance.
2. **Model Selection and Development:** Explore a variety of machine and deep learning algorithms, including decision trees, support vector machines, neural networks, and convolutional neural networks, selecting the most suitable models for the given task.
3. **Hyperparameter Tuning and Optimization:** Fine-tune model parameters to enhance performance and generalization capabilities, ensuring robustness against overfitting.
4. **Evaluation and Interpretation:** Rigorously evaluate model performance using appropriate metrics, visualize results, and interpret model decisions to gain insights into underlying patterns.
5. **Real-World Application:** Demonstrate the practical applicability of the developed models by deploying them on real-world data and showcasing their potential impact.

Through the successful execution of these objectives, this project endeavours to contribute to the advancement of classification techniques in machine learning, paving the way for innovative solutions that address complex challenges and unlock new possibilities.

Dataset Overview and Preprocessing:

Describe the datasets used for each task, the preprocessing steps taken, and any challenges encountered during this stage.

* The datasets for both tasks were vastly different.
* Dataset one,
* Dataset two, was a zip file in which included a wide range of images that had digits from 0-9. Preprocessing this particular dataset was difficult. Due to the large amounts of images in different folders, it was easier to upload the file into GitHub ready for the CNN model.

### Task 1: Numerical and Categorical Classification

* 1. **Methodology and Techniques**: Explain the machine learning models and techniques applied to the first task, including model selection, hyperparameter tuning and evaluation metrics.
  2. **Results and Discussion**: Present the results, performance metrics, and data visualisations for the first task, and discuss the implications of your findings.

### Task 2: Multi-label Image-based Digit Classification

* 1. **Methodology and Techniques**: Explain the machine learning models and techniques applied to the second task, including model selection, hyperparameter tuning and evaluation metrics.
* For this particular classification, the technique used was Convolutional Neural Network model (CNN). The data provided was vast amounts of images of digits; a CNN model is suited to handling image variability and scaling to large datasets.
* The key reasons for choosing a CNN model is that a CNN model provides an in-depth analysis for multi-image-based digit classification.
  1. Results and Discussion: Present the results, performance metrics, and data visualisations for the second task, and discuss the implications of your findings.

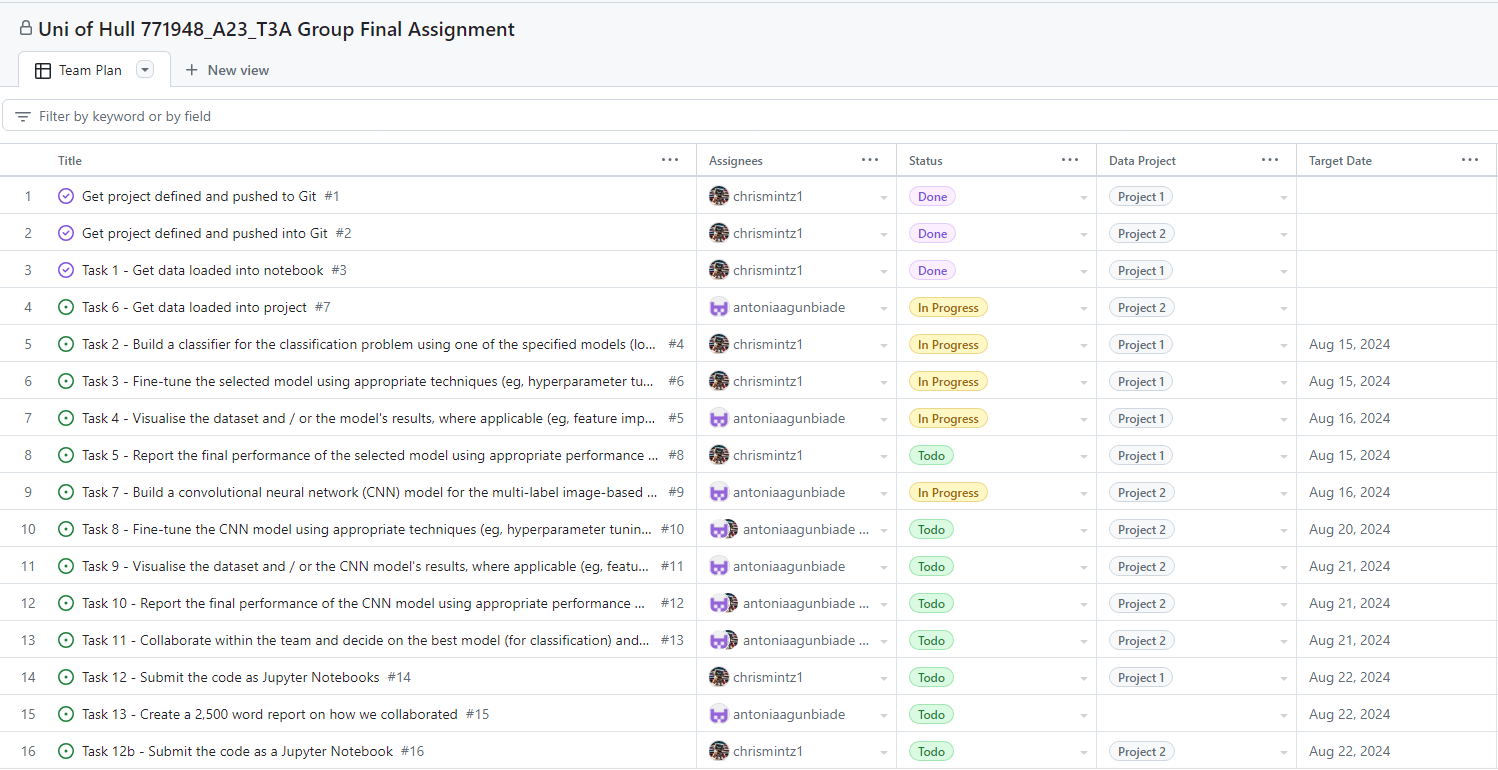
Model Comparison and Selection: Compare the performance of the models developed for both tasks, discuss the trade-offs between the models, and explain your final model selections.

Conclusion: Summarise the key insights from the assignment, the implications of your findings, and provide suggestions for future work or improvements.

## Collaboration:

The collaboration for this project, was initiated through WhatsApp. A messaging platform that allowed the team members to get acquainted with each other as well as the project at hand. By initially contacting each other via WhatsApp, it allowed for an icebreaker in order to understand each team members’ working style, strengths and weaknesses to be able to find a suitable way to complete the project. The only obstacle faced during the initial phases of the project was scheduling meetings in line with the different time- zones of the team members. However, this was quickly resolved, and regular Teams and WhatsApp meetings were scheduled in order to track progress of the project.

This then led to allocation of tasks, meeting scheduling required for the kick off the project.

GitHub, was then the platform decided to collaborate on for the project. Allowing for the creation of models and etc.. Through GitHub, a project was created in which tasks were allocated to each member of team, along with the status of task and target date for completion. As shown in the screenshot above.

Branches were used to work on tasks then rebased and merged to main when the team felt ready to share their updates.